AMENDMENTS

Amendments to the Specification

Please replace the paragraph on page 1, lines 7-14 with the following amended paragraph:

This application is related to U.S. Patent Application Serial Number 09/721,042, filed on November 21, 2000, entitled "Methods and Computer Software Products for Predicting Nucleic Acid Hybridization Affinity"; U.S. Patent Application Serial Number 09/718,295, filed on November[[,]] 21, 2000, entitled "Methods and Computer Software Products for Selecting Nucleic Acid Probes" and U.S. Patent Application Serial Number ~~_/____, attorney docket number 3373.1~~ 09/745,965, filed on [[___]] December 21, 2000, entitled "~~Methods For Selecting Nucleic Acid Probes~~ Method and Software Products for Selecting Probes Using Dynamic Programming." All the cited applications are incorporated herein by reference in their entireties for all purposes.

Please replace the paragraph on page 7, lines 7-14 with the following amended paragraph:

Microarrays ~~Microarray~~ can be used in a variety of ways. A preferred microarray contains nucleic acids and is used to analyze nucleic acid samples. Typically, a nucleic acid sample is prepared from an appropriate source and labeled with a signal moiety, such as a fluorescent label. The sample is hybridized with the array under appropriate conditions. The arrays are washed or otherwise processed to remove non-hybridized sample nucleic acids. The hybridization is then evaluated by detecting the distribution of the label on the chip. The distribution of label may be detected by scanning the arrays to determine fluorescence intensity distribution. Typically, the hybridization of each probe is reflected by several pixel intensities. The raw intensity data may be stored in a gray scale pixel intensity file. The GATC™ Consortium has specified several file formats for storing array intensity data. The final software specification is available at the GATC Consortium's website ~~www.gatcconsortium.org and is incorporated herein by reference~~

~~in its entirety~~. The pixel intensity files are usually large. For example, a GATC™ compatible image file may be approximately 50 Mb if there are about 5000 pixels on each of the horizontal and vertical axes and if a two byte integer is used for every pixel intensity. The pixels may be grouped into cells (see, GATC™ software specification). The probes in a cell are designed to have the same sequence (i.e., each cell is a probe area). A CEL file contains the statistics of a cell, e.g., the 75th percentile and standard deviation of intensities of pixels in a cell. The 50, 60, 70, 75 or 80th percentile of pixel intensity of a cell is often used as the intensity of the cell.

Please replace the paragraph on page 11, lines 17-29 with the following amended paragraph:

FIGURE 3 shows an exemplary computer network that is suitable for executing the computer software of the invention. A computer workstation 302 is connected with the application/data server(s) through a local area network (LAN) 301, such as an Ethernet 305. A printer 304 may be connected directly to the workstation or to the Ethernet 305. The LAN may be connected to a wide area network (WAN), such as the Internet 308, via a gateway server 307 which may also serve as a firewall between the WAN 308 and the LAN 305. In preferred embodiments, the workstation may communicate with outside data sources, such as the National Centre for Biotechnology Information (NCBI) ~~Center~~, through the Internet. Various protocols, such as FTP and HTTP, may be used for data communication between the workstation and the outside data sources. Outside genetic data sources, such as the GenBank 310, are well known to those skilled in the art. An overview of GenBank and ~~the National Center for Biotechnology information (~~NCBI~~)~~ can be found ~~in~~ on the web site of NCBI ~~(http://www.ncbi.nlm.nih.gov)~~.

Please replace the paragraph on page 12, lines 16-25 with the following amended paragraph:

FIGURE 4 shows an exemplary process for designing a gene expression probe array. Sequences from various sources are used for sequence selection 401. The source sequences may come from, *e.g.*, genomic sequences, cDNA sequences, expressed sequence tags (ESTs) or EST clusters. The sequence selection process generates candidate sequences for probe selection 402. For photolithographic synthesis of oligonucleotide arrays, masks may be designed based upon the probe sequences 403. Processes, systems and computer software products for probe selection and mask designs are disclosed in, for example, U.S. Pat. Nos. 5,800,992, 6,040,138, 5,571,639, 5,593,839, and 5,856,101, and U.S. Patent Application Serial Nos. 09/719,295, 09/721,042 and 09/745,965 ~~Attorney Docket Number 3273.1~~, all incorporated herein by reference for all purposes.

Please replace the paragraphs on page 13, lines 9-12, 13-21 and 22-28 with the following amended paragraphs:

The sequence selection process may involve the use of clustering tools, BLAST ~~(http://www.ncbi.nlm.nih.gov)~~, FASTA, etc. In addition, gene identification/prediction tools, multiple alignment tools, consensus calling/assembly methods may also be employed.

FIGURE 5 shows an exemplary process for sequence selection for expression probe arrays. Raw sequence information from public or private databases, mRNA, Coding Regions (CDS), EST, gene clusters (such as UniGene Clusters) or genomic sequences, from various public or private databases, such as Genbank, UniGene, *etc.* are cleaned 501. For a review of genetic databases, see, e.g. ~~*see, e.g.*~~, Searls (2000). Bioinformatics Tools For Whole Genomes. *Annu. Rev. Genom. Hum. Genet.* 1: 251-279, which is incorporated herein by reference for all purposes. A catalog of genetic databases, called Biocatalog, is available ~~at~~ on the EMBL-EBI (European Molecular Biology Laboratory – European Bioinformatics Institute) website

~~http://www.ebi.ac.uk/biocat/, last visited on January 11, 2000, the content of the web site is incorporated herein by reference for all purposes~~.

Cleaning sequences may be as late as into the probe design phase in at least some embodiments. However, in preferred embodiments, the cleaning is performed as early as possible, *i.e.*, in the first stage of the entire sequence selection phase, particularly if UniGene sequences (~~http://www.ncbi.nlm.nih.gov~~ available on the NCBI website) are used as input to the sequence selection. The output of the cleaning process 501 is a set of cleaned EST/mRNA sequences. Cleaning is often necessary because, even though UniGene had screened its sequences against ribosomal RNAs, vector contamination, and low-complexity regions, a large numbers of UniGene sequences with repetitive elements, low complexity regions, and ambiguous regions still exist.

Please replace the paragraph on page 23, lines 22-26 with the following amended paragraph:

In preferred embodiments, a simple computational method is used to derive *p*. The method includes obtaining binomial frequency distribution of subclusters of size n over the number of contradictory sequences consistent with D_{minor}. ~~*P*~~ *p* may be estimated using $p' = f^1f_{max}(x)/n$, where $f^1f_{max}(x)$ is the value of x when f(x), or b(x; n, *p*), reaches its maximum.